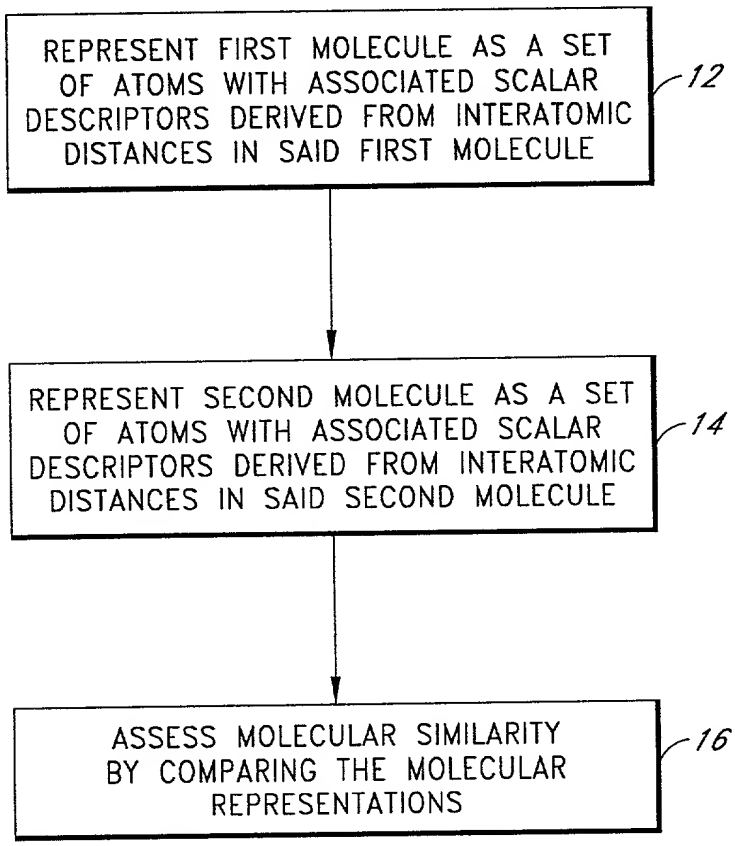

REPRESENT FIRST MOLECULE AS A SET OF ATOMS WITH ASSOCIATED SCALAR DESCRIPTORS DERIVED FROM INTERATOMIC DISTANCES IN SAID FIRST MOLECULE
12
REPRESENT SECOND MOLECULE AS A SET OF ATOMS WITH ASSOCIATED SCALAR DESCRIPTORS DERIVED FROM INTERATOMIC DISTANCES IN SAID SECOND MOLECULE
14
ASSESS MOLECULAR SIMILARITY BY COMPARING THE MOLECULAR REPRESENTATIONS
16

FIG. 1

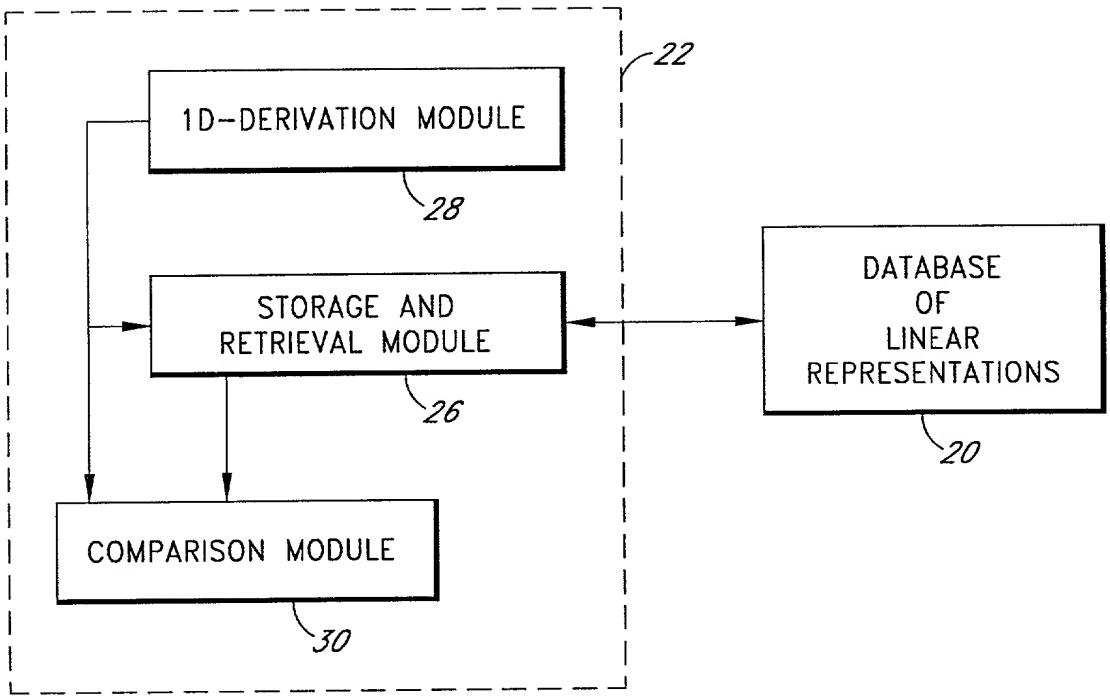

22
1D-DERIVATION MODULE
28
STORAGE AND
RETRIEVAL MODULE
26
COMPARISON MODULE
30
DATABASE
OF
LINEAR
REPRESENTATIONS
20

FIG. 2

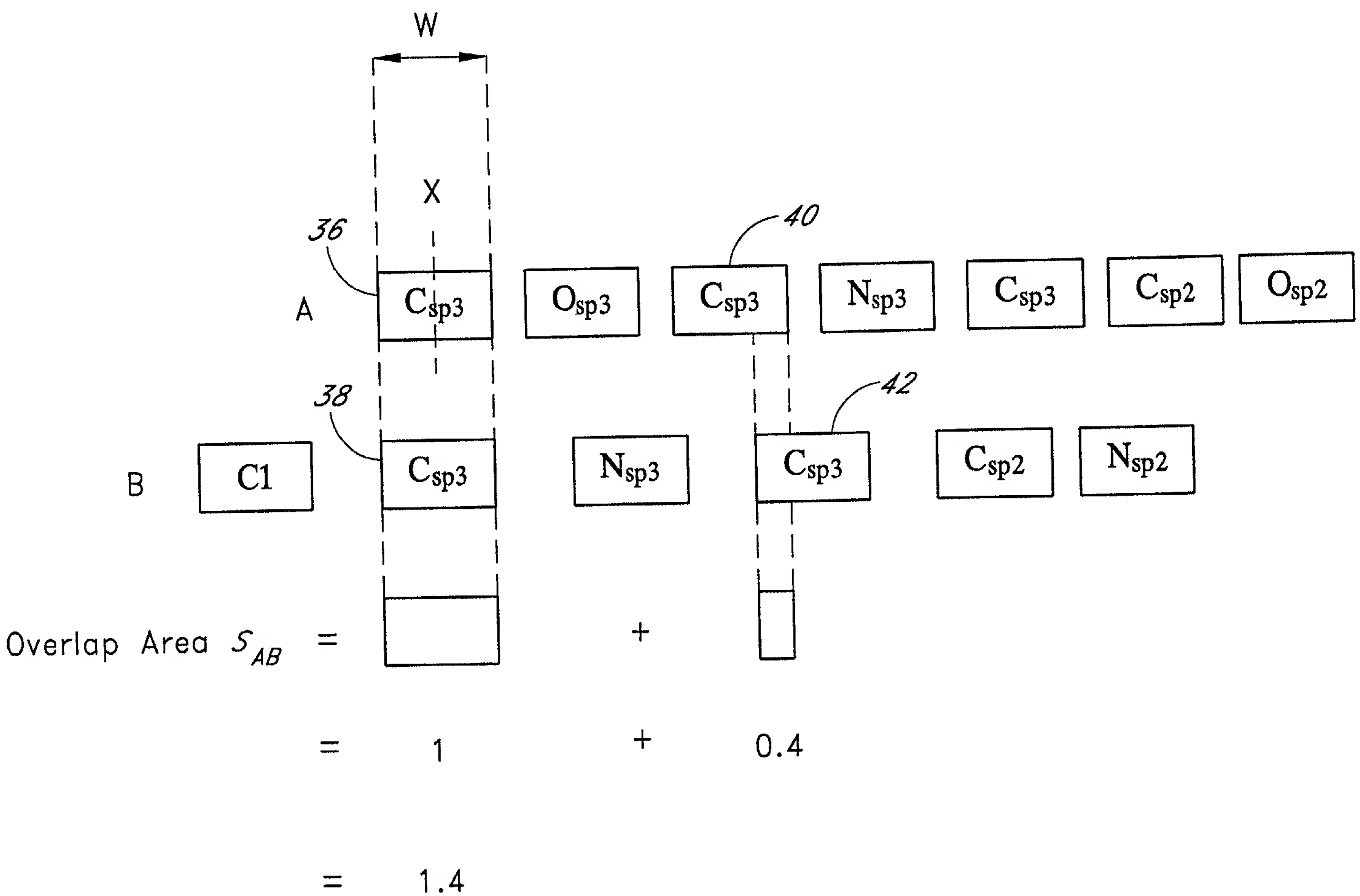
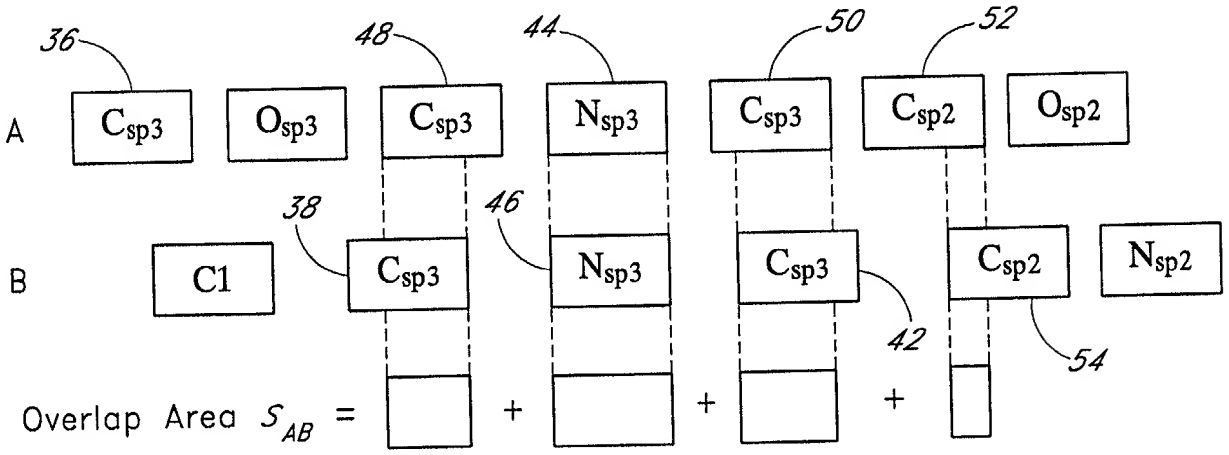

W
X
36
40
38
42
A
B
C1
C_{sp3}
O_{sp3}
C_{sp3}
N_{sp3}
C_{sp3}
C_{sp2}
O_{sp2}
C_{sp3}
N_{sp3}
C_{sp3}
C_{sp2}
N_{sp2}
Overlap Area S_{AB} =
+
= 1 + 0.4
= 1.4

FIG. 3A



= 0.6 + 1 + 0.8 + 0.4

= 2.8

FIG. 3B

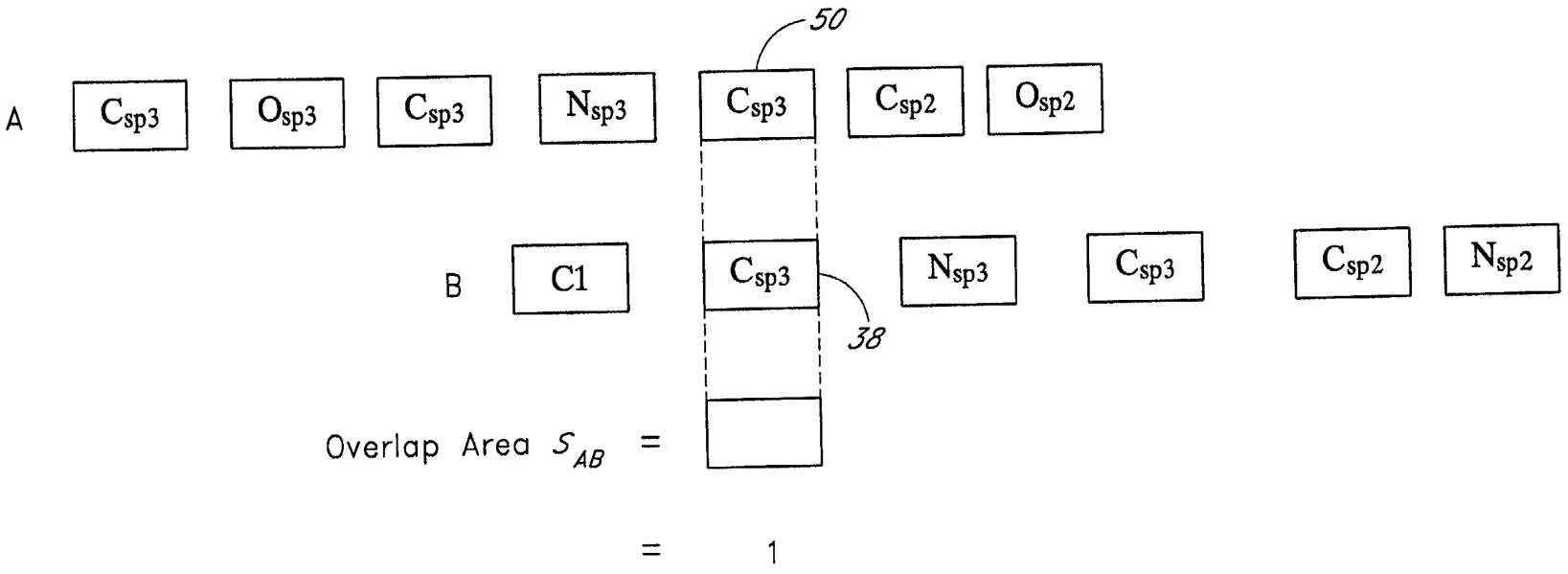


FIG. 3C

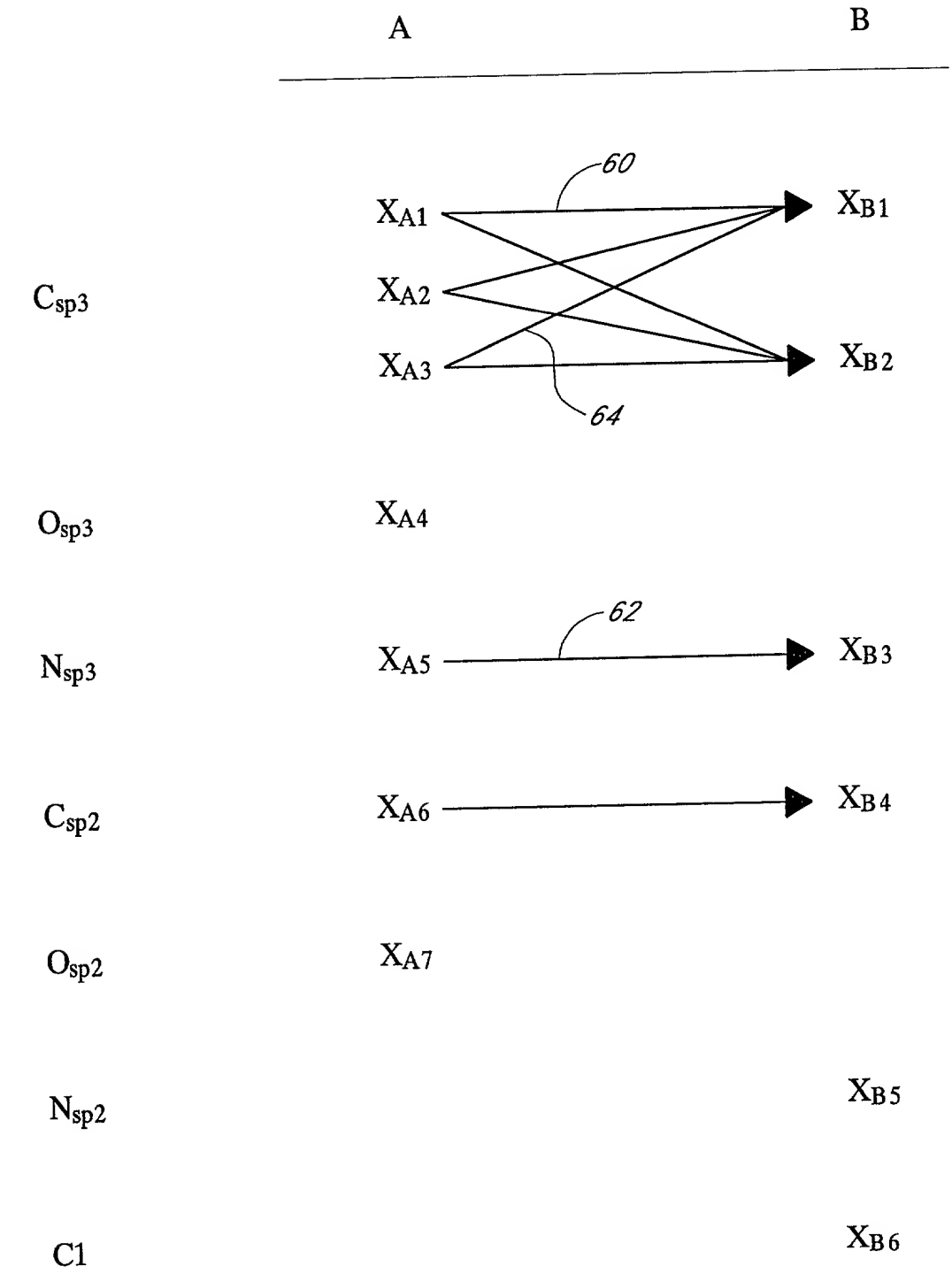

A
B
60
X_{A1}
X_{A2}
X_{A3}
X_{B1}
X_{B2}
64
C_{sp3}
O_{sp3}
X_{A4}
N_{sp3}
62
X_{A5}
X_{B3}
C_{sp2}
X_{A6}
X_{B4}
O_{sp2}
X_{A7}
N_{sp2}
X_{B5}
C1
X_{B6}

FIG. 4

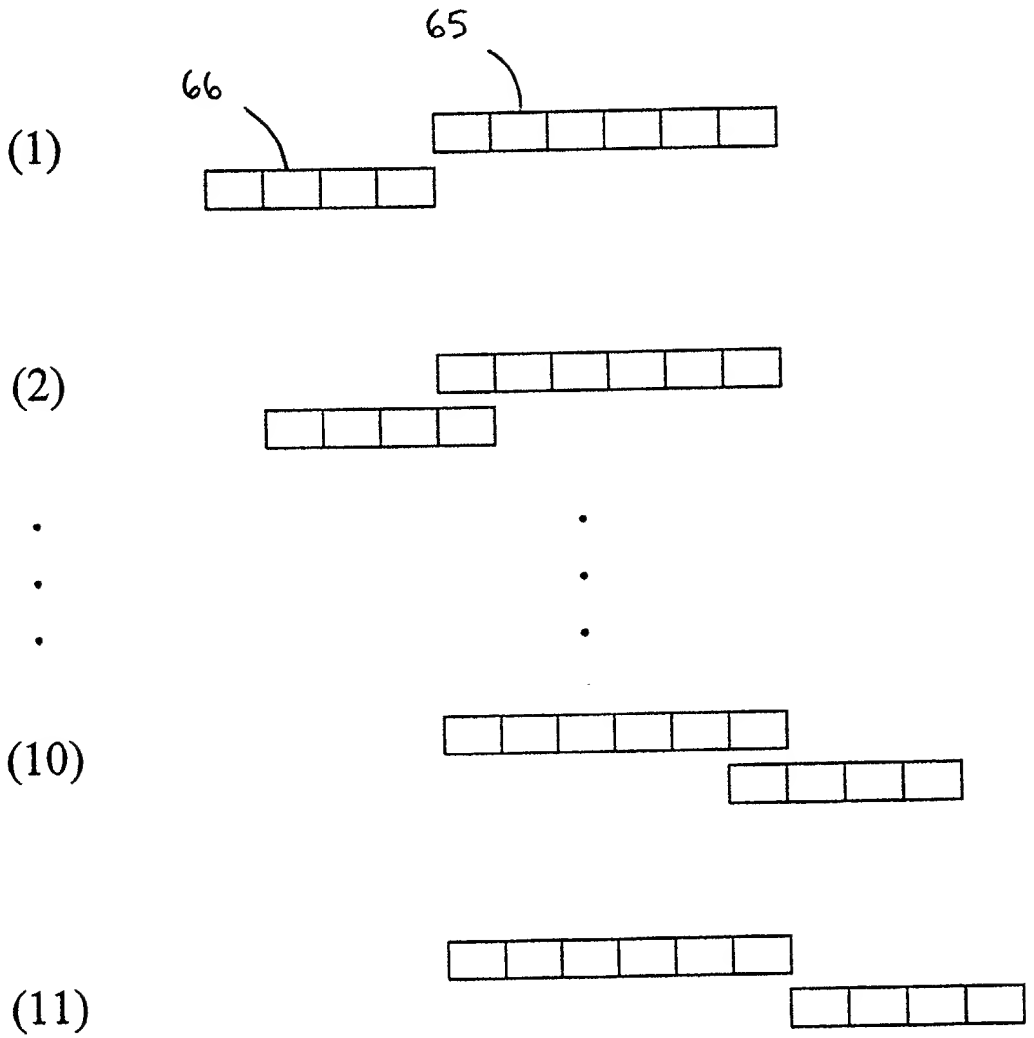


FIG. 5

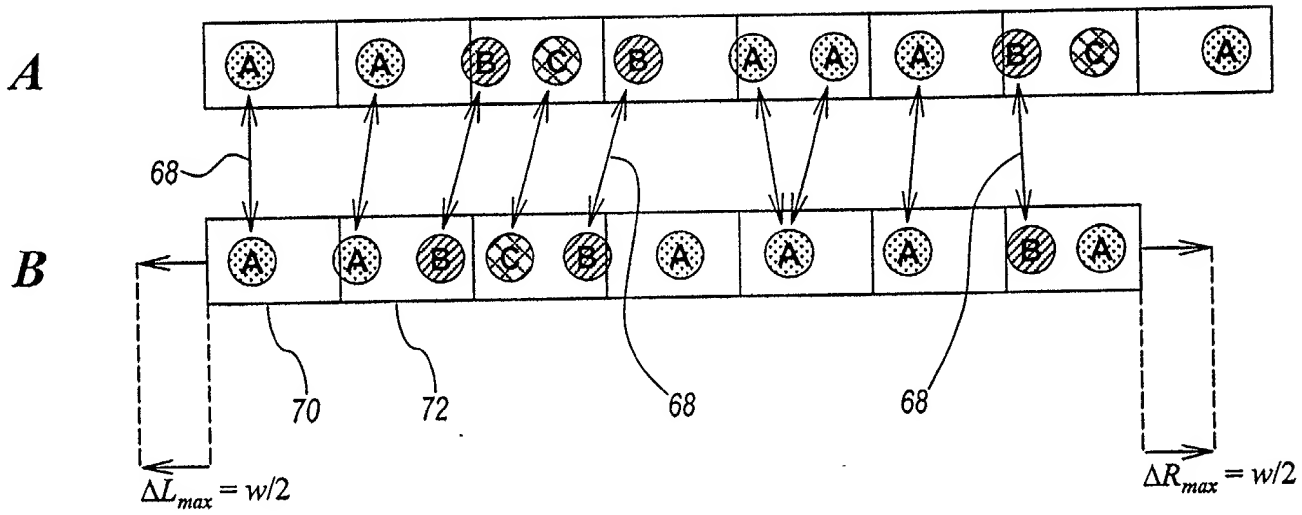


FIG. 6

FIG. 7A

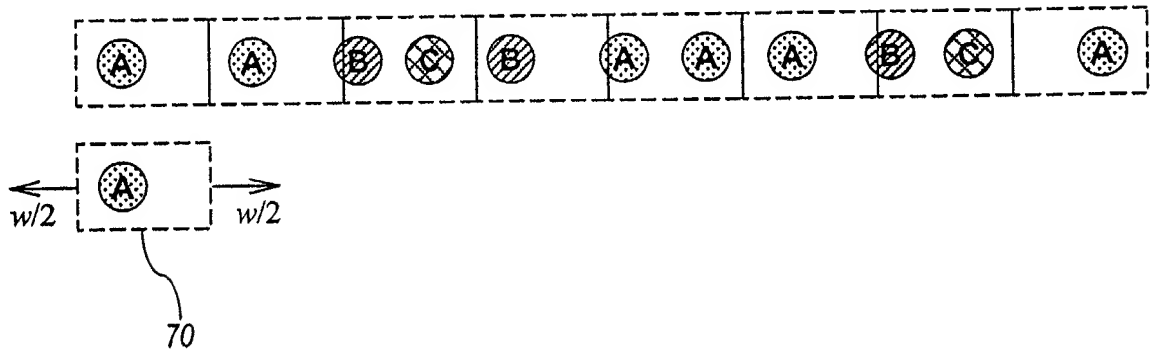


FIG. 7B

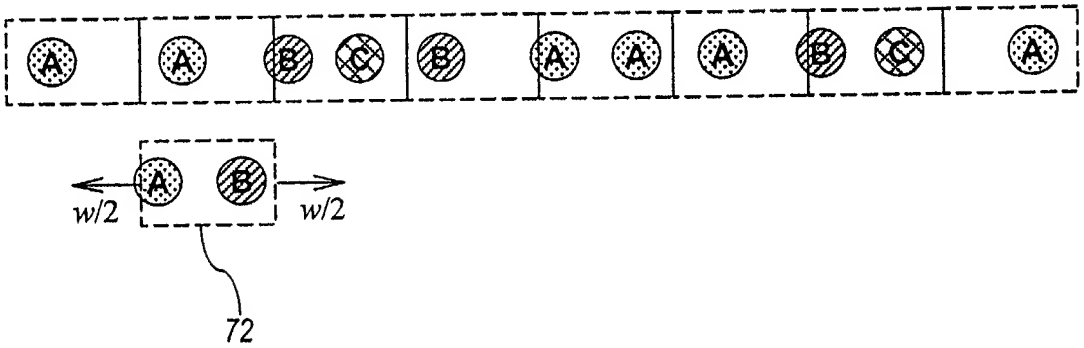
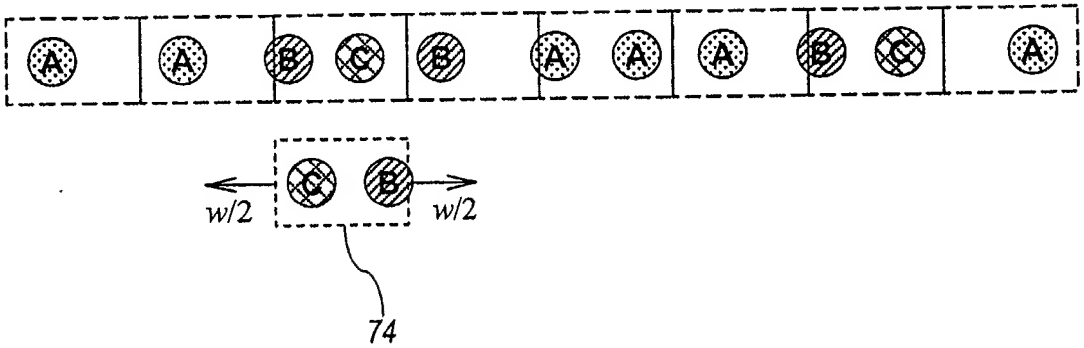


FIG. 7C



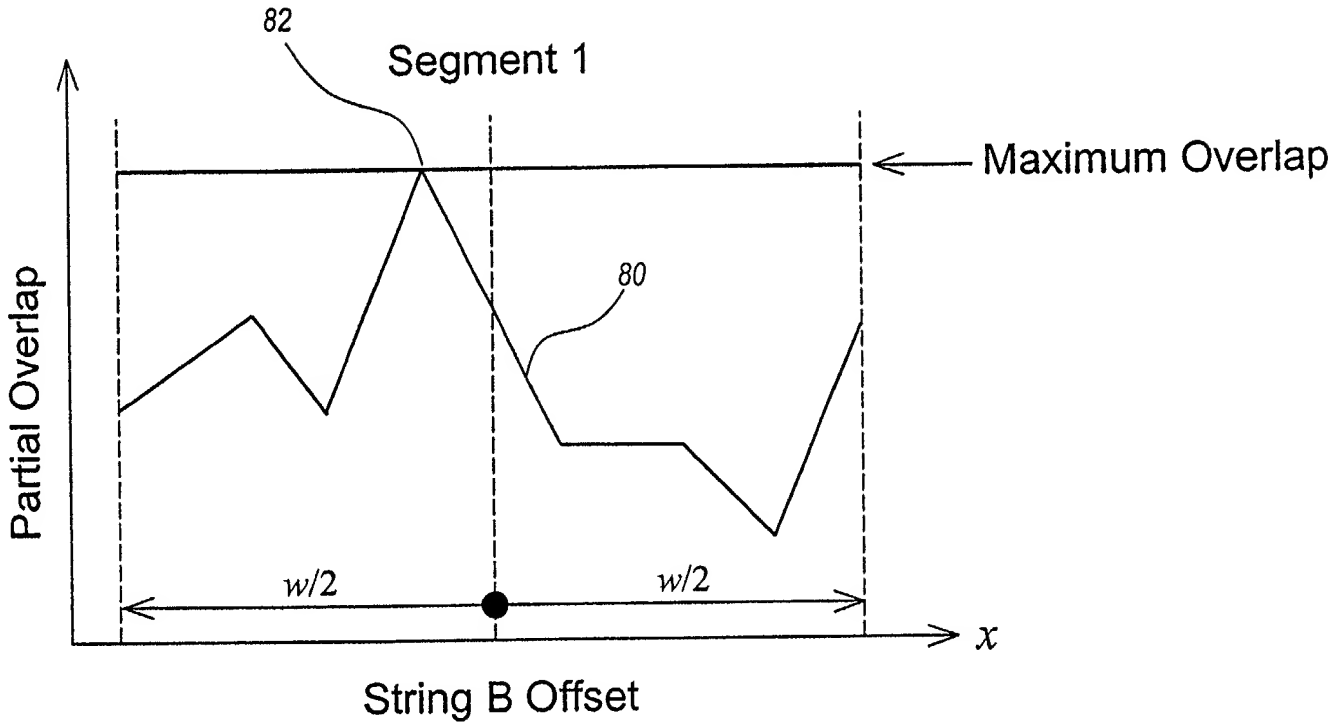


FIG. 8A

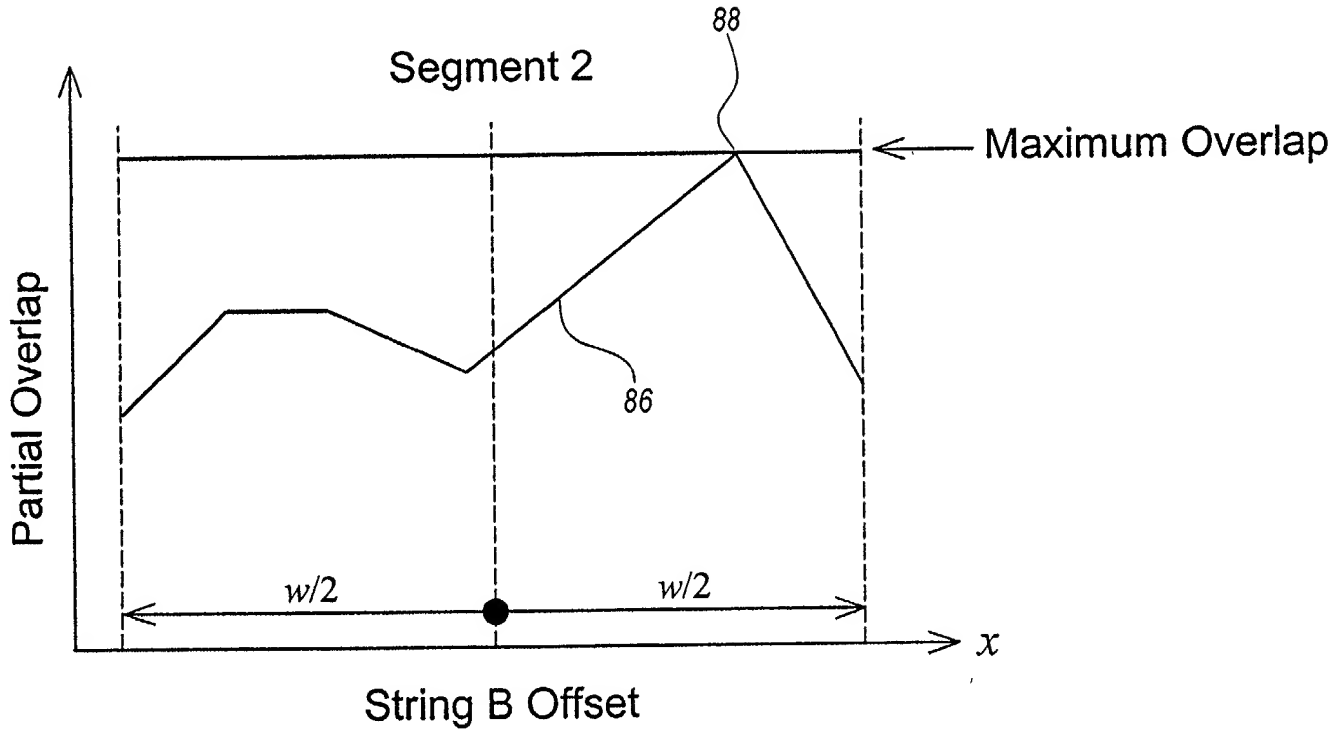


FIG. 8B

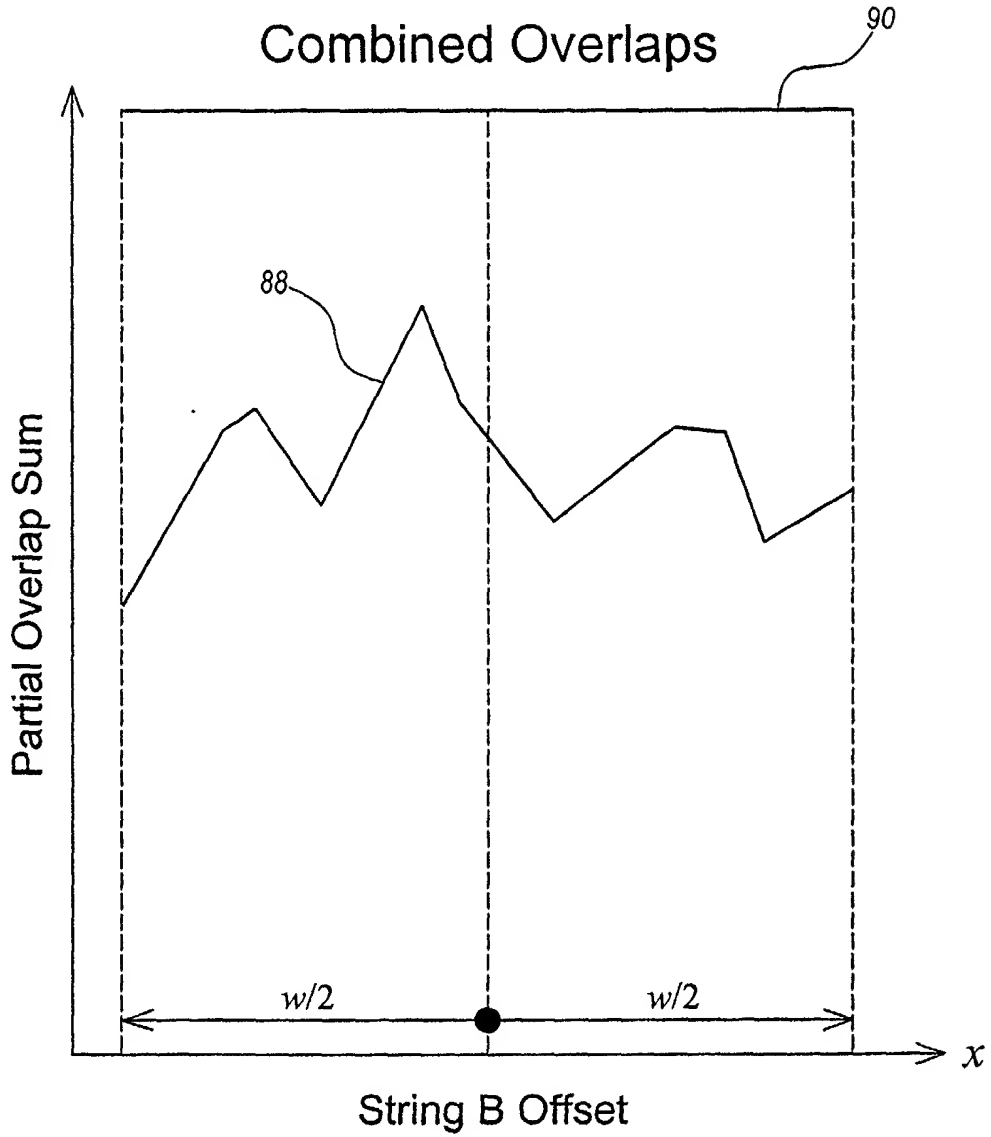


FIG. 8C

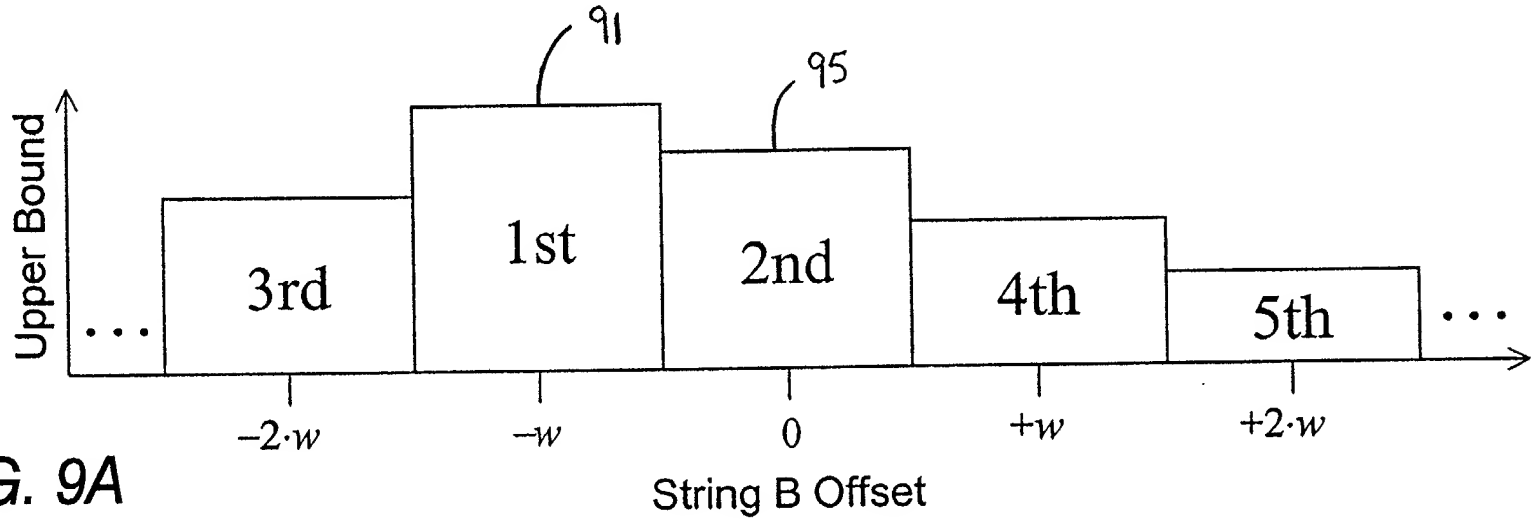

91
95
Upper Bound
3rd
1st
2nd
4th
5th
...
...
−2·w
−w
0
+w
+2·w
String B Offset

FIG. 9A

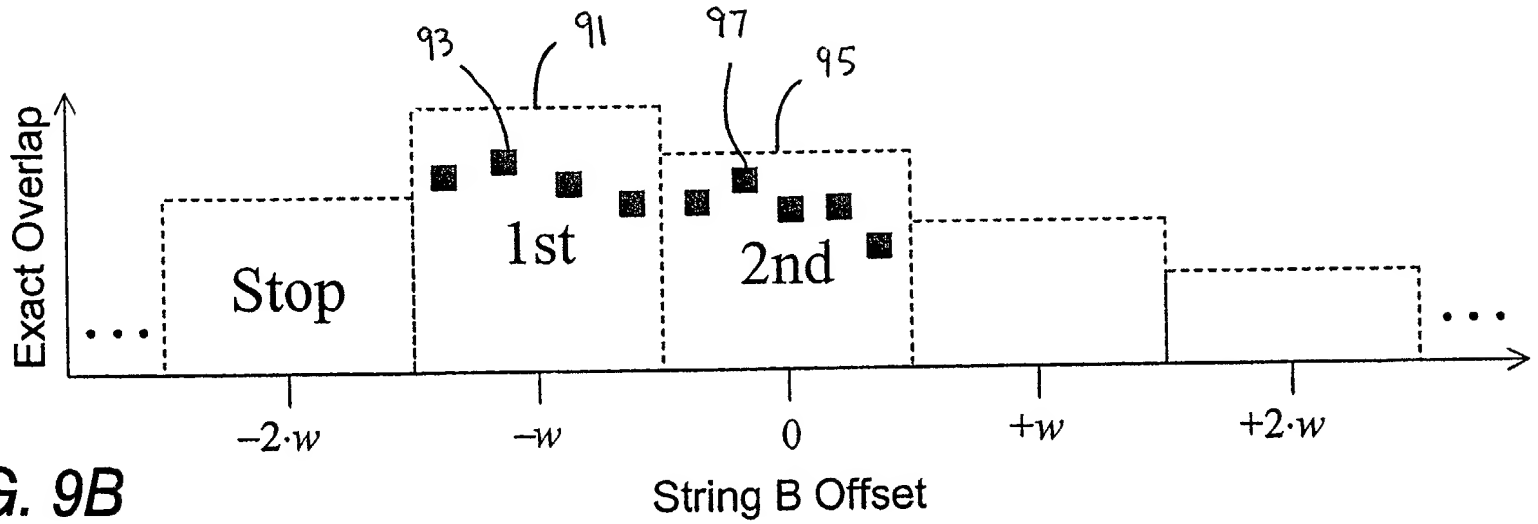

93
91
97
95
Exact Overlap
Stop
1st
2nd
...
...
−2·w
−w
0
+w
+2·w
String B Offset

FIG. 9B

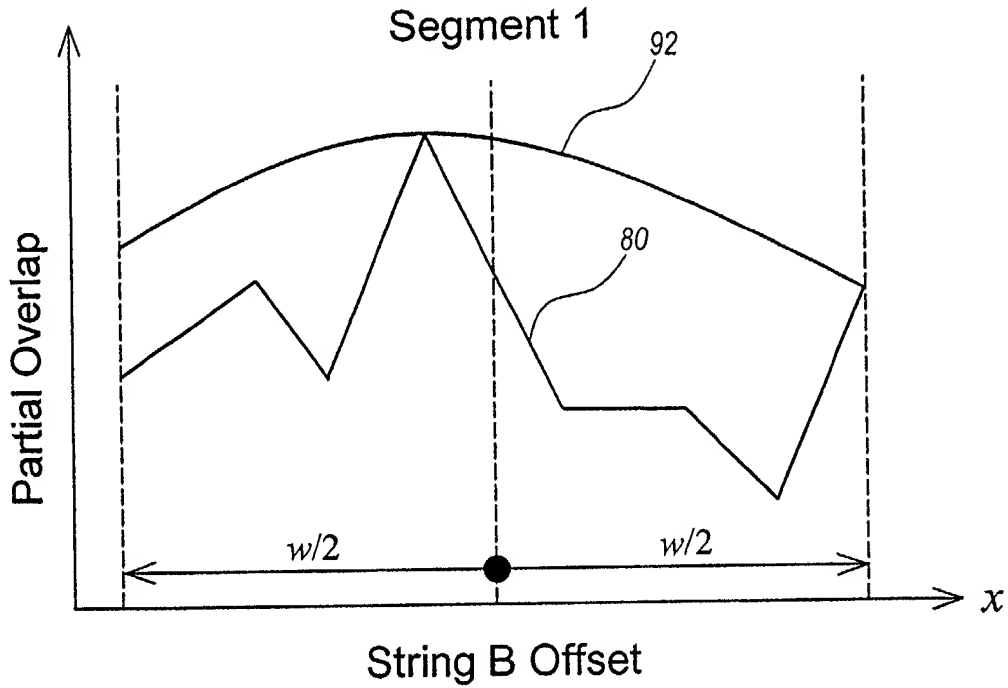


FIG. 10A

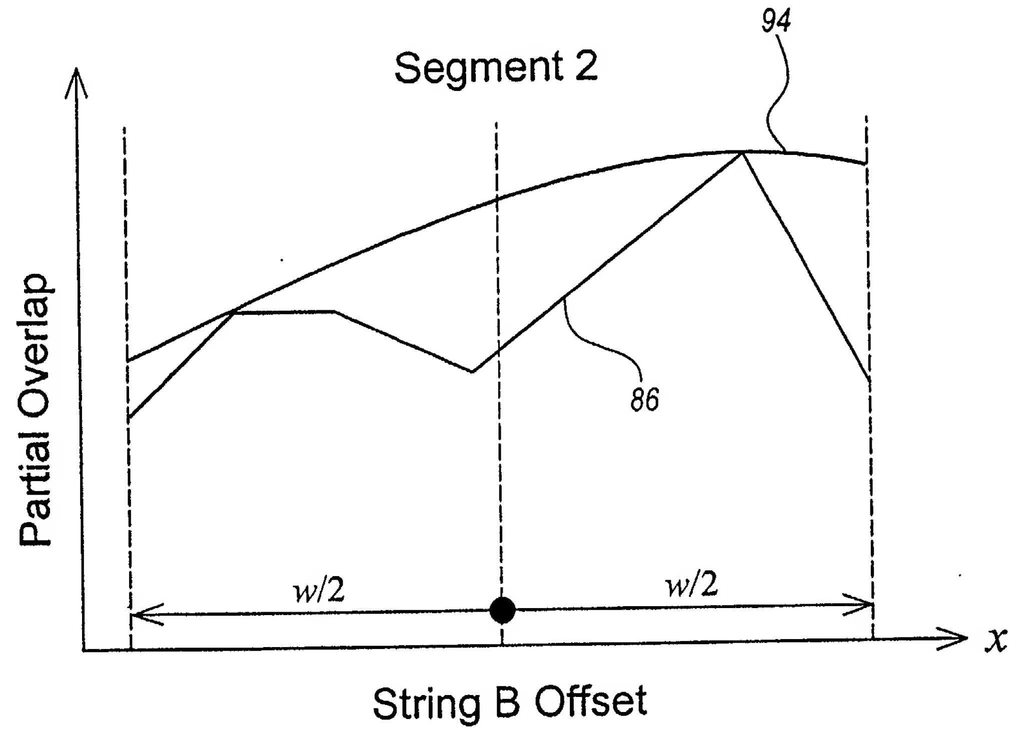


FIG. 10B

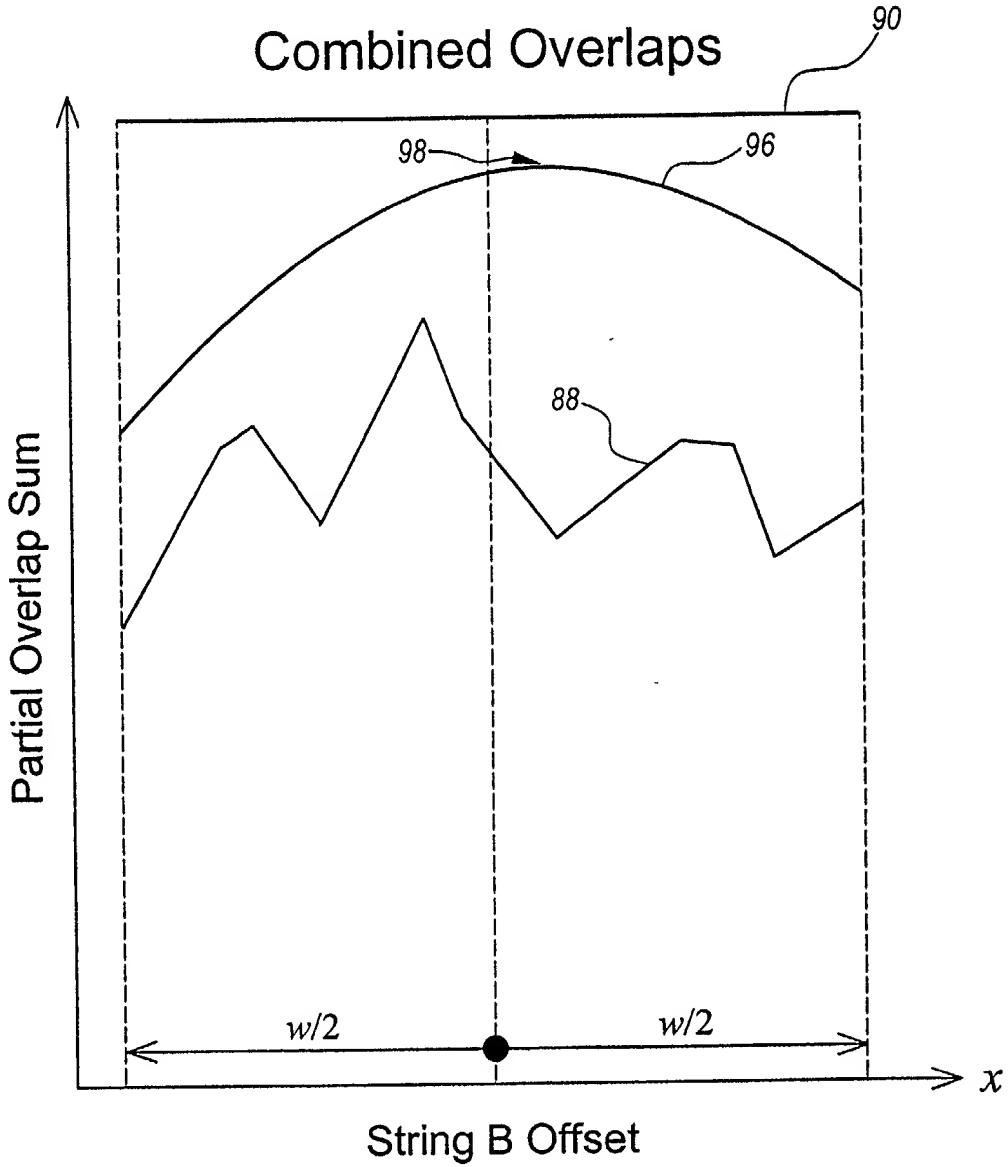
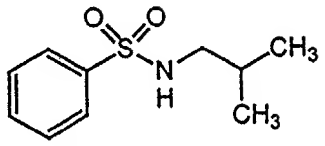
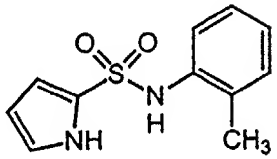


FIG. 10C

3D→1D

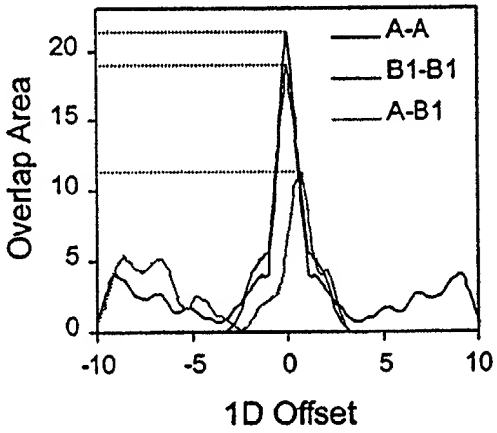


Compound A



Compound B1

2D→1D



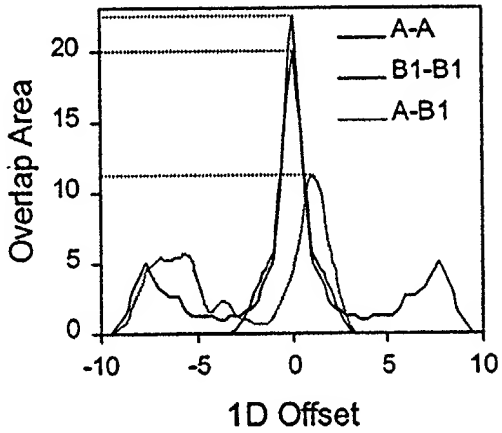
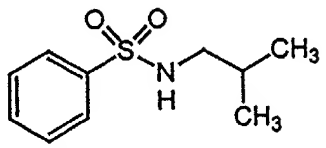
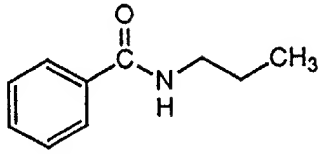


FIG. 11

3D→1D

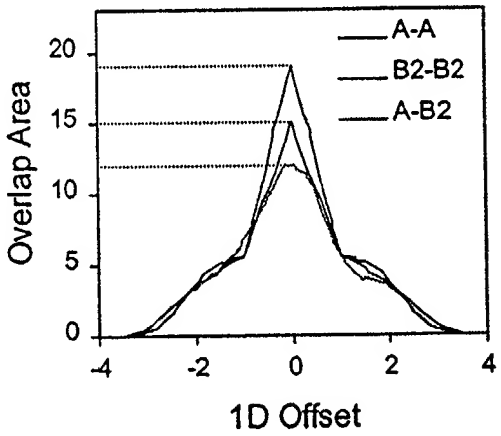


Compound A



Compound B2

2D→1D



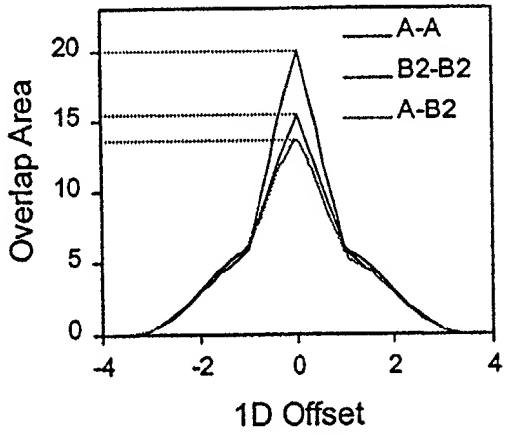
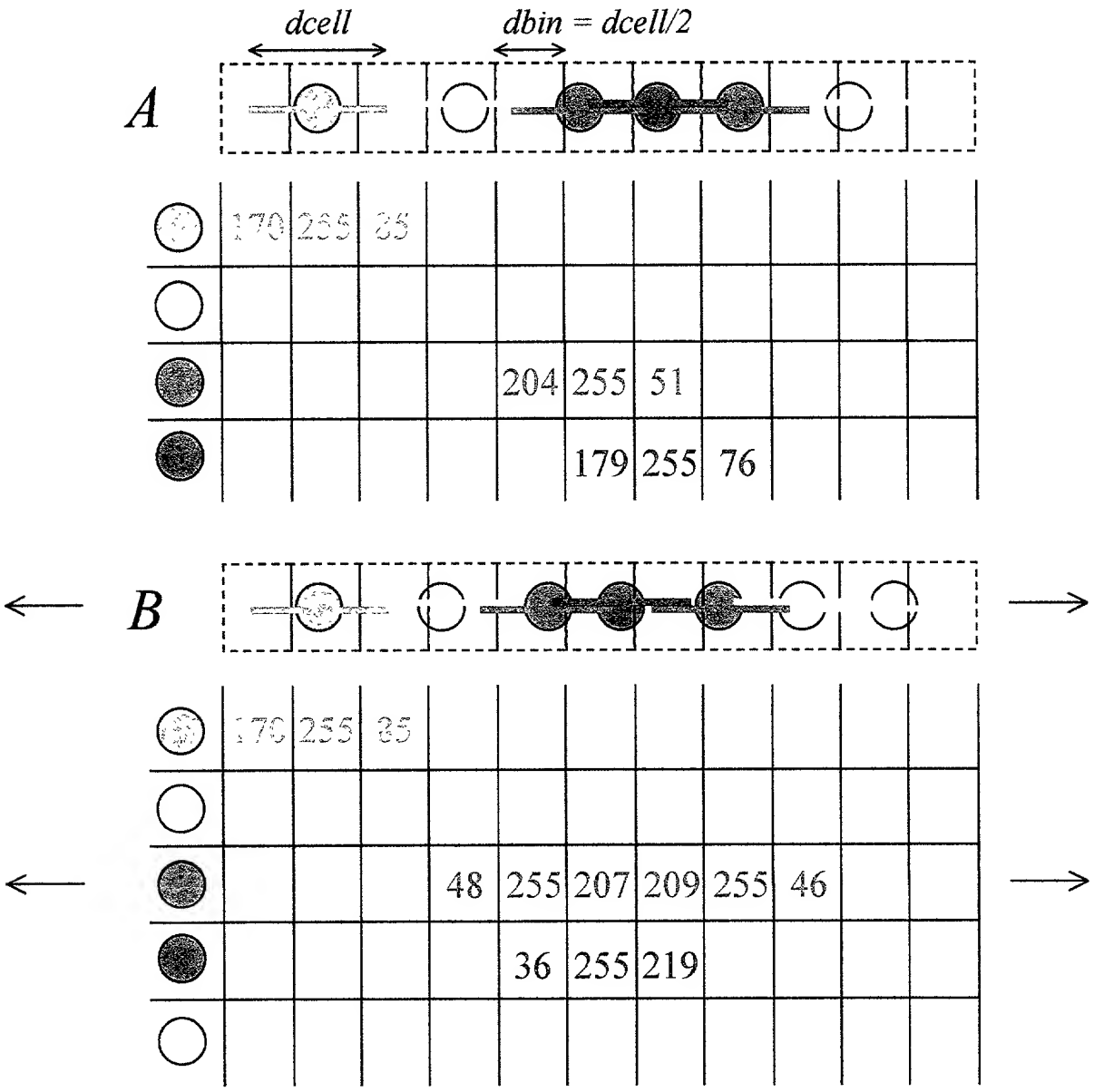


FIG. 12

Bin-Based Overlap

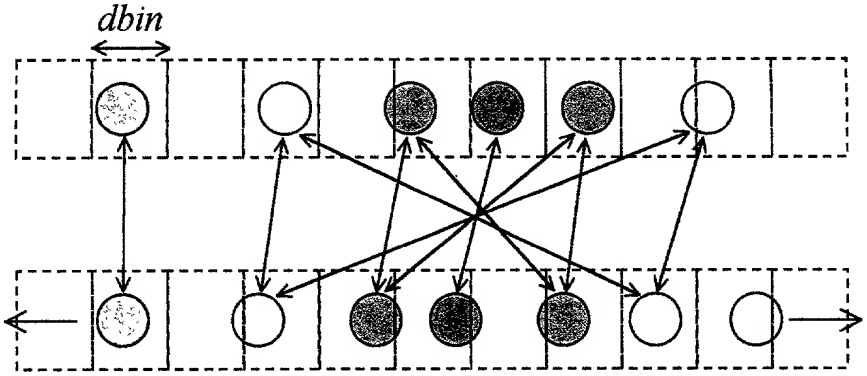
- Do a series of fast overlap calculations using "bins" with integer occupation numbers (0→255) for each atom:



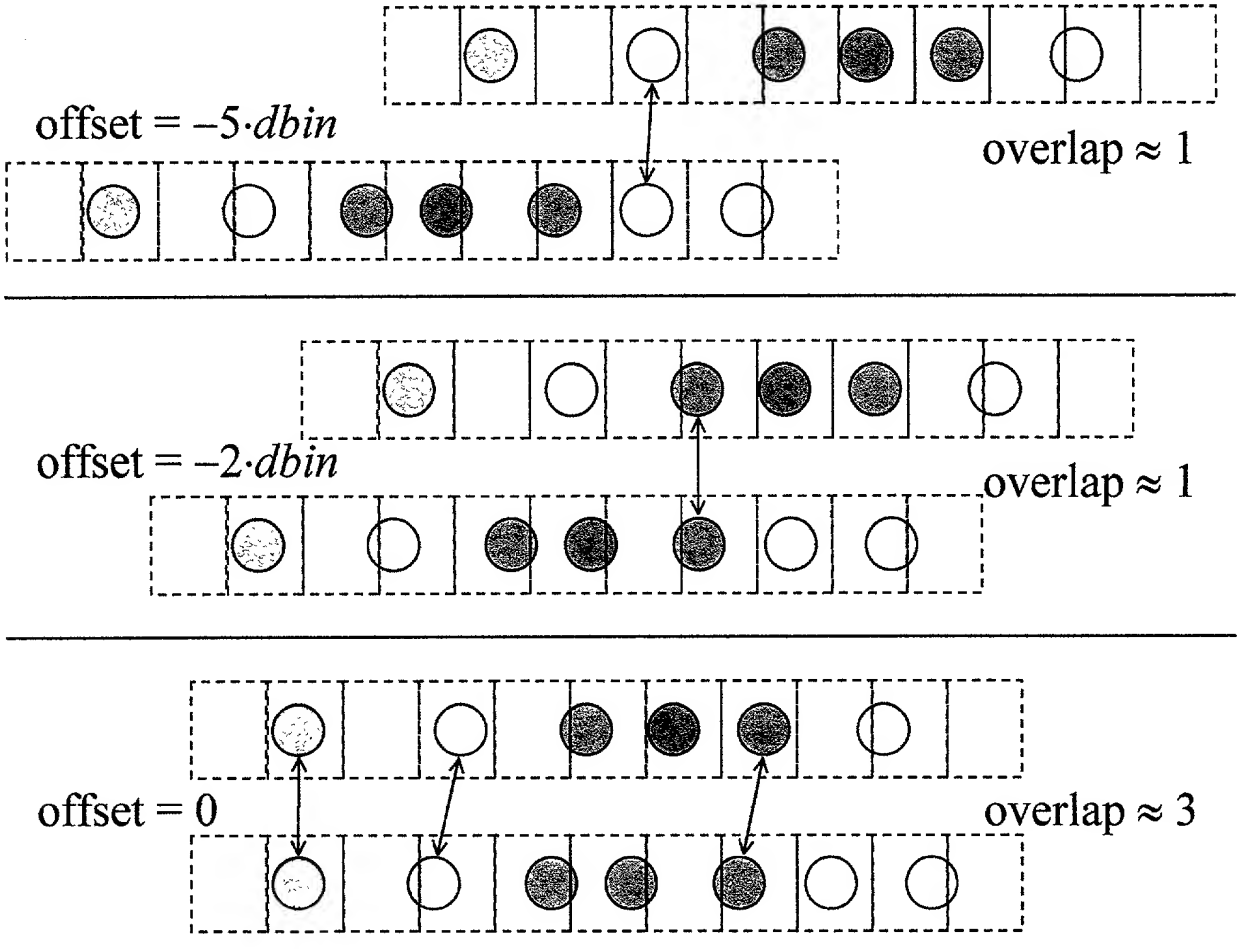
- Multiply occupation numbers for matching atom types across aligned bins to get a good estimate of overlap area
- Fast, but there are numerous bin-based offsets that must be considered

FIG. 13

Speeding Up Bin-Based Overlap Calculations



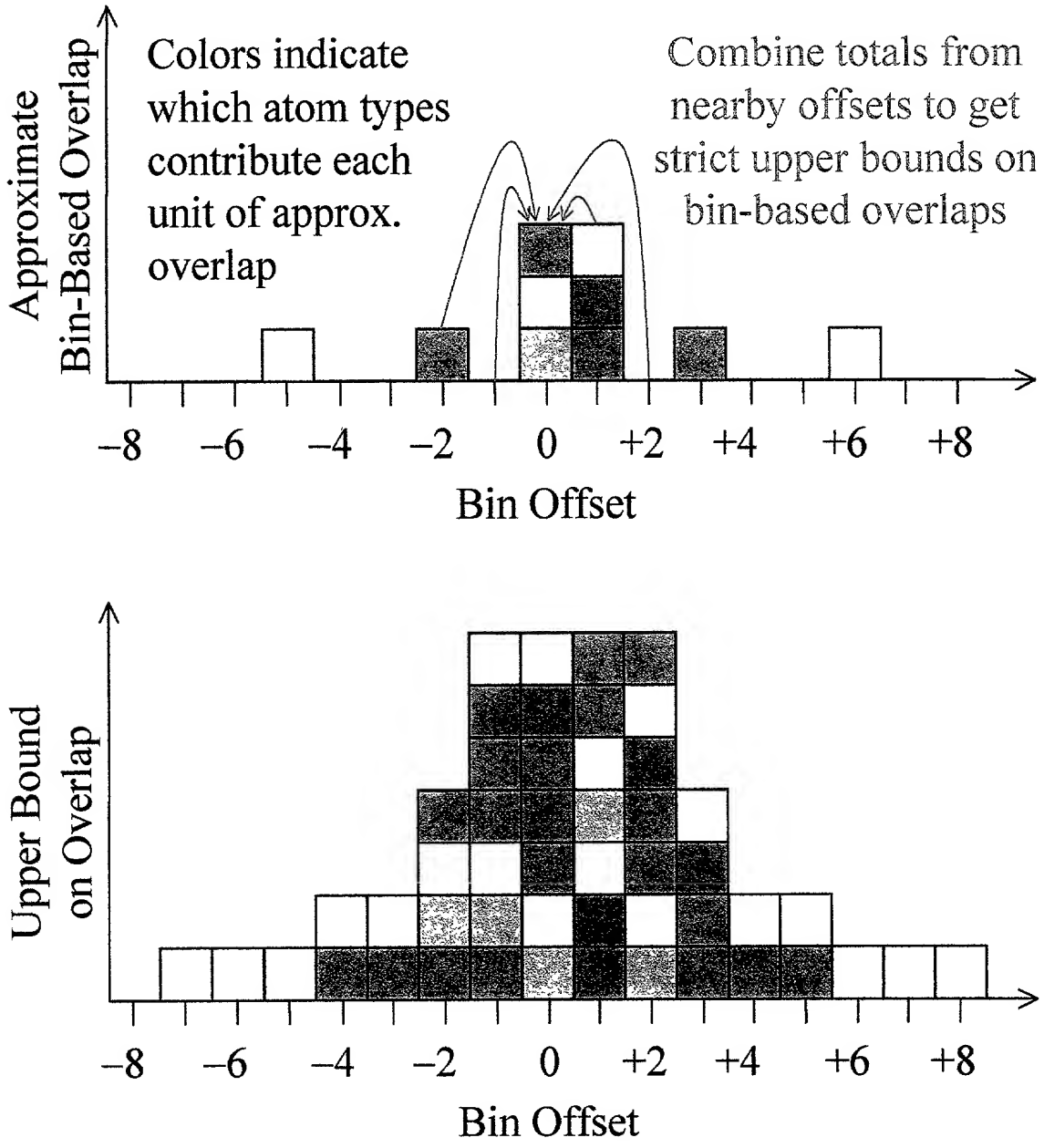
- 21 unique bin offsets, 10 matching atom type pairs
- There are only 6 different bin offsets wherein matching atom types are approximately aligned:



etc.

FIG. 14

Approximate Bin-Based Overlaps → Upper Bounds



- Process offsets in order of decreasing upper bound
- Do standard bin-based overlap calculations (with occupation numbers), keeping track of the largest overlap value
- Stop when remaining upper bounds are lower than this largest bin-based overlap

FIG. 15